# Peer Interaction and Scientific Reasoning Processes in Preschoolers

Guevara Guerrero, Marlenny



*Publication date:*
2015

# Chapter 5

## The Complexity of Children's Reasoning: Actions, Verbalizations and Transfer in Spontaneous Experimentation

# The Complexity of Children's Reasoning: Actions, Verbalizations and Transfer in Spontaneous Experimentation[9]


## Abstract

This study examines how the processes of reasoning and transfer evolve in young school children during a series of problem solving tasks. First, we examined the complexity of reasoning as two different qualitative processes: implicit and explicit knowledge. Secondly, we examined transfer as a history-dependent process which occurs from each task to the next. In a context of spontaneous experimentation, seven dyads of 5-year-olds ($M_{age}$= 5.1 years) participated in this study. The dyads were asked to solve a set of hands-on tasks, increasing in complexity over six sessions. The results indicate that on average the complexity of children's reasoning was higher for actions than for verbalizations, both within sessions and over time. This means that children identified more structural relations of the underlying mechanism in their actions than in their explanations. The intra-individual analysis reveals that children showed variable patterns of (implicit and explicit) reasoning and transfer. These results suggest that the children's trajectories of reasoning and transfer vary and self-organize over time as a result of the intricate relation with the affordances of the tasks.


*Keywords*: explicit knowledge, implicit knowledge, reasoning, skill levels, transfer, children, variability, nonlinearity

---

[9] This chapter is based on Guevara, M., van Dijk, M., & van Geert, P. (2015). *The Complexity of Children's Reasoning: Actions, Verbalizations and Transfer in Spontaneous Experimentation*. Manuscript in preparation.

## 1. Introduction

Promoting reasoning skills is central in the educational agenda of many modern societies. The two most important aspects concern the complexity of reasoning and its transfer across contexts (Sjøberg, 2003). These two aspects are often stimulated in educational settings by providing learning contexts in which learners can build knowledge and apply it to new situations, inside and outside the classroom. The development of scientific reasoning skills is often seen as an essential feature of becoming a scientifically literate adult (Croke & Buchanan, 2011) who is able to explain and predict natural phenomena, evaluate arguments based on evidence (NRCl, 1996), and apply conclusions from such arguments. Most science education interventions aim to nurture, enrich and sustain children's natural and spontaneous interest in scientific knowledge and procedures (Klahr, Zimmerman & Jirout, 2011) and to enhance understanding of scientific phenomena (De Lange, et. al. 2008; Kuhn, 2011). In addition, a modern society that demands innovations, requires people to adapt to changes and face new situations by transferring knowledge (see Vries, 2003). For this reason, educational systems invest in preparing students for future learning processes through the use of important skills such inquiry (Brandsford & Schwartz, 1999) and the integration of implicit and explicit knowledge (Sun, Slusarz & Terry, 2005). The impact of education is measured by the way in which learners are able to transfer their knowledge to other contexts than those in which a particular skill or knowledge has been taught.

A specific difficulty faced by educators is the low achievement of transfer, a phenomenon also known as 'inert knowledge' (Perkins, 1992), in which learners tend to isolate pieces of information (Hale, 2013). It seems that what people have learned at a particular moment is not available or accessible at another moment (Larsen-Freeman, 2013). Some authors argue that this may be mainly a result of task-specific instructions, the specific domains of knowledge and the type of knowledge that is to be transferred (Butterfield & Nelson, 1991). However, there is a lack of literature addressing the emergence of spontaneous transfer over time. In the present study we provide a combination of a microgenetic and a longitudinal approach to explore the development of reasoning in young school-age children and their transfer of scientific reasoning during a series of problem-solving tasks. Knowledge about the emergence of these processes is necessary in order to improve our understanding of the nature of scientific reasoning in young children. Accordingly, we need more insights about the actual temporal patterns of scientific reasoning and transfer over the short and long-term timescale (For temporal analyses of interaction see also Chapters 1 and 2 of this dissertation).

### 1.1. Scientific Reasoning in Young Children

Scientific reasoning can be understood as the application of methods of inquiry to a problem-solving situation (Kuhn & Franklin, 2007). According to the Scientific Discovery as Dual Search (SDDS) model (Klahr, Fay & Dunbar, 1993), inquiry skills refer to three main domain-general scientific reasoning processes: hypothesis-generation, experimentation and evidence evaluation. Based on the average of age groups, Piekny and Maehler (2013) found an asynchronous development of these three skills during early and middle childhood (6-10 year olds). The development of evaluating evidence starts by understanding that experimenting is different from producing a desired outcome. The development of this skill is observed during the transition from preschool to the first primary school years. After this, the hypothesis-generation skill appears, being the most difficult for young children.

Studies on scientific reasoning have not only explored the use of different skills of reasoning in verbalizations (explicit knowledge) but also in actions (implicit knowledge) (see Beilock & Goldin-Meadow, 2010; Garnham & Perner, 2001). Earlier studies (e.g. Goldin-Meadow (2010, 2014) have indicated that gestures and actions convey information that not always is present in verbalizations. Despite the consensus on the distinction between implicit (procedural or *know-how*) and explicit knowledge (declarative or *know-what*), there is much debate about how particular knowledge develops over time. One perspective, proposed by Anderson (1983), corresponds to the theory of skill acquisition as an adaptive control of thought (ACT). This theory suggests that knowledge emerges as a top-down process: first as explicit knowledge, which is automatized with practice and later becomes implicit for memory efficiency. In contrast, a bottom-up approach suggests that implicit knowledge develops before explicit knowledge: children and adults “can perform a task with increasing awareness and efficiency before they are able to completely understand and explain their success” (Piekny & Maehler, 2013, p.285). For instance, Ebersbach and Resing (2008) found that there are more five-year-olds who can use magnitudes to depict linear and exponential growth than there are five-year-olds who can explain the change of these quantifications over time. Similarly, Pine and Messer (1999) showed that the relation between implicit and explicit knowledge is age-related. They found that 5-year-olds show a good performance in a balance scale task but poor verbalizations explaining their outcomes, while 9-year-olds succeed in both.

Another example in which implicit knowledge precedes explicit knowledge is the case of gestures. Previous studies (Goldin-Meadow, 2003; Goldin-Meadow & Alibali, 2013) have shown that children first appear to display an idea in a gesture before they are able to express it verbally. Also, these nonverbal performances represented a higher cognitive level than their verbalizations. In addition, it is argued that when a child is asked to explain a problem-solving task, there is no way of ensuring that the child's verbalized explanation is a reliable indicator of what the child knows (Pine, Lufkin & Messer, 2004). An alternative approach argues that implicit and explicit knowledge are interconnected and develop simultaneously (Sun, Merrill, & Peterson, 2001; Willingham & Goedert-Eschmann, 1999).

The literature on this subject shows that from early childhood onwards, children are able to use causal models as a mechanism of learning (Carey, 1995; Gopnik, Sobel, Schulz & Glymour, 2001; Goswani, 1991; Waismeyer, Meltzoff & Gopnik, 2014). Children are able to construct causal theories that help them make sense of their experiences, even though these theories are often incorrect and incomplete (Kuhn, 2011). In everyday activities, causal knowledge can be seen as a spontaneous skill, which takes place during exploratory learning activities and play. In these activities, children manipulate objects, and perform spontaneous experimentation. Several studies support the idea that children are able to elaborate causal models to predict and design interventions (Buchsbaum, Bridgers, Weisberg and Gopnik, 2012; Cook, Goodman & Schulz, 2011).

In general, mature scientific reasoning requires, among others, two fundamental things. One is the ability to demonstrate implicit and explicit reasoning, and the other is the ability to establish causal relations. Despite the number of studies about reasoning, little is known about the individual development of these abilities at the early school age, on both, the short- and long-term time scale. Though some longitudinal designs have been used to study reasoning and the ability to transfer reasoning skills, many are based on a single or a limited number of measurement points which are separated by long intervals (Brandsford & Schwartz, 1999). This does not provide information about which are the patterns of change and how the complexity of reasoning evolve over time. The approach of the present study is to focus on the process of emergence of scientific reasoning and the ability to transfer – by using repeated measures– and the group and intra-individual differences in the short- and long- term patterns.

## 1.2. Dynamic Systems Theory and the Complexity of Reasoning

The complex dynamic systems approach to scientific reasoning is explicitly interested in the individual process of development and the eventual inter-individual differences in those temporal patterns. This approach argues that development of reasoning is the result of a process of self-organization that takes place in the form of a continuous person-environment loop. The interrelation of the individual's behavior in the context is defined as a process of 'soft-assembly'. In this process the components of behavior can interact depending on the circumstances of the context in which they take place (Kloos & van Orden, 2009). In this sense, soft assembly contrasts with the view that reasoning is a direct reflection of an internal pattern of action (often called an internal representation). Instead, it is seen as a temporal integration of a wide variety of internal and external contextual conditions assembled 'on the spot'. From this view, the context is not something that acts like an external co-determinant of the way the ability is expressed, but an integral and a constituent part of the ability as a process of real-time performance.

In addition, the complex dynamic systems approach claims that scientific reasoning has an iterative nature, which indicates co-causality. For instance, a behavior caused at time point *t+1* is co-determined by the previous behavior at time point *t*. Every time children interact in a context, they use their previous knowledge and preceding experiences (i.e. the preceding performance). Because of a constant feedback loop with the environment, small differences between the children's initial understanding can increase or decrease, depending on whether the feedback is positive or negative (van der Steen, Steenbeek & van Geert, 2012).

According to the complex dynamic systems approach, variability is another central characteristic of development and not a random fluctuation based on external independent conditions (van Dijk & van Geert, 2015). Therefore, intra-individual variability shows a characteristic structure (e.g. Kello, Anderson, Holden & van Orden, 2008). For instance, several studies have shown that an increase in variability can precede a developmental transition (see Bassano & van Geert, 2007; Lewis & Cook 2007; van Dijk & van Geert, 2007). In the course of a reasoning task, the performance of children can show intra-individual variability (van der Steen, Steenbeek & van Geert, 2012) and this is currently seen as an important factor that may be related to the ability to reach higher levels of reasoning or to a transition to another pattern of behavior (Howe & Lewis, 2005). For instance, Kloos, Fischer and van Orden (2010), have demonstrated that variability in cognitive development is closely related to the interdependence between the individual and the task constraints. Smith and Thelen

(2003), state that behaviors are variable and context-dependent, suggesting that they need to be studied and understood as systems, which "always are in flux" (p. 343). This means that the complexity of the skills displayed in a task is related to changes in task context or content. More specifically, Vallacher, Nowak & van Geert (2014) indicate that variability is a property of the *intrinsic* dynamics of behavioral processes instead of a result of an extrinsic source of the context. With regard to this, van Dijk and van Geert (2015) point out that variability plays an important function in the process of development. On the one hand, it provides information about the transitions when cognitive change takes place. On the other hand, variability is a functional component of development, as the force and cause of change. For these reasons, the authors argue that variability is important when studying developmental change.

Fischer's dynamic skill theory (Fischer 1980) offers a framework to explain the variability of the cognitive development based on the principles of dynamic systems theory. In this theory (Fischer, 1980), the complexity of scientific reasoning is described through three tiers: sensorimotor (2-5 years), representational (4- 8 years), representational system (7-12 years) and abstractions (12-20 years). Each tier is made up of three levels: 'single', 'mapping' and 'system' (Fischer & Bidell, 2006). At the *sensorimotor* tier, the child's focus lies on the physical sensorimotor experiences, for instance on perceiving and describing the object itself. Later, on the sensorimotor mapping level, the child combines two sensorimotor characteristics, such as the object and its attributes. At the system level, two mappings can be combined and manipulated as an observed result and a related action. In the second tier of single *representations*, the child coordinates several systems and can understand that an object has a characteristic beyond the situation in which it has been observed. At the mapping level, several characteristics can be linked together as establishing causal relations. It is not until the final level of the representational tier that a child can relate two mappings in one more complex and detailed skill. The third tier corresponds to *abstractions*, where the child generalizes ideas about an event beyond the concrete observed situation, as principles. Skill Theory also provides a model for the transition among complexity levels of scientific reasoning, which in our case can be considered as a reference for the transition between implicit and explicit knowledge and the complexity levels of the ability to transfer.

In relation to these complexity levels, Bidell and Fischer (1994) state that people's capacities are not fixed but can vary across a range of levels at any time during development. Because there is much variability in each child's activities, the levels of reasoning that are described are not meant as distinct stages, but as a description of the quantitative change in skill organization. For these authors, the variability of the cognitive levels is related to the cognitive demand that underlies the relation between individual and context. For example, on a *functional level*, children's skills are the result of their spontaneous performance on a task without practice or support. On an *optimal level*, the children's skills are the result of their upper level of understanding, as a result of practice and/or scaffolding.

Skill theory also describes variability of the cognitive trajectories as an important feature of development and describes a "scalloping" pattern in which periods of high variability and low variability alternate (Fischer, 2008). Thus, the process of learning is conceived of as a dynamic cycle in which higher levels of understanding are the result of a repeated process of building a skill.

### 1.3. The Nature of Transfer

Performances vary because the psychological structures underlying them are the result of the self-organizing process that takes place in real-time performance as well as on the long-term level of development (Fischer & Bidell, 2006). Hence, we expect that transfer is also variable in a similar way. Transfer should therefore not be conceptualized as a mechanic 'exportation' of knowledge, but as a transformation of it (Larsen-Freeman, 2013; Nokes, 2009). Nokes (2009) suggests that learners can use at least three mechanisms to transfer knowledge according to the demands of the task and the knowledge that needs to be transferred. One is the most common mechanism of transfer, based on the analogical relations between tasks or their structural similarities (see Gentner, & Markman, 2006). The other two mechanisms correspond to 'compilation', in which learners translate declarative knowledge in procedures, and 'constraint violation', that refers to an ongoing process of 'generate-evaluate-revise' actions to solve the task. Another important aspect of transfer is that the specific affordances of the different tasks can increase the chances that transfer takes place (Day & Goldstone, 2012). For example, in order to transfer what has been learned in situation A to situation B, it is necessary that the learner identify the structural similarities and differences between tasks –, adjusting their actions in function of the new goal.

Looking for an understanding of transfer as a dynamic process, we can complement the previously mentioned mechanisms of transfer with the concepts of 'self-organization' (i.e. the order or coordination emerges from the local interactions between the components) and 'soft–assembly' (i.e. individual-context interaction) (see van der Steen, Steenbeek & van Geert, 2012). We argue that transfer is the result of soft assembly of performance that is self-organizing instead of being a mechanical move of knowledge from one domain to another. During transfer, experiences and memories resulting from earlier problem-solving situations are becoming integrated —soft-assembled — in a similar or new domain.

### 1.4. Research Questions

In the previous section, we have introduced the three main aspects of our study. First, scientific reasoning during spontaneous experimentation consists of implicit (e.g. actions to solve the tasks) and explicit (e.g. verbal explanations indicating the causal relation of the tasks) knowledge (Klahr, Fay, & Dunbar, 1993). Second, from a complex dynamic systems perspective, reasoning skills are conceived of as emerging from the interaction of the individual with the context, i.e. as 'soft-assembled' phenomena (Thelen & Smith, 1994; Spencer, Perone & Johnson, 2009; van der Steen, Steenbeek & van Geert, 2012). Third, there is a need to rethink transfer as a dynamic process, in which learning is transformed and adapted in function of the demands of the context. Adopting a complex dynamic systems approach emphasizes the importance of studying individual patterns of development. Hence, for a better understanding of scientific reasoning and transfer we must also have knowledge about the individual patterns of their development on the short and long-term timescale.

This study is part of a research program called *Curious Minds* (TalentenKracht in Dutch, www.talentenkracht.nl) in which researchers at universities in The Netherlands and Belgium elaborate on the study of diverse aspects of children's skills in science and technology, with the goal of promoting children's talent and providing means to improve educational practices in the areas of science and technology. In the current study, we propose two aims. First, we aim to investigate the short-term as well as long-term emergence of the complexity of reasoning of individual children as implicit knowledge (action events, AE) and explicit knowledge (verbal events, VE) by using the framework of skill levels as a quantitative measure of such emergence. We work with 5-year-olds, because according to Fischer and others (1980; Fischer & Bidell 1998) these children are expected to function around the transition between giving sensorimotor (single representations) to representational system (representational mapping) explanations. The second aim is to study the emergence of transfer as an integral aspect of the development of scientific reasoning in the context of spontaneous experimentation with hands-on task sets about physics. Specifically, the research questions are:

1. How does children's complexity of reasoning as implicit (AE) and explicit knowledge (VE) emerge over time?
2. How comparable are the intra-individual trajectories of the levels of complexity of children's implicit (AE) and explicit knowledge (VE), in two different task sets?
3. Do the two task sets elicit similar complexity levels of reasoning for implicit (AE) and explicit knowledge (VE)?
4. How does transfer of scientific reasoning emerge in the children's AE from task to task?

## 2. Method

### 2.1. Participants

Participants were fourteen 5 year olds ($M_{age}$= 5.1 years, SD= 0.48; 7 girls and 7 boys) who were part of a larger sample of a study on reasoning and peer interaction. Participants were recruited from an international school in The Netherlands and all of them were proficient in English. The children worked in dyads that were brought together by their teacher, based on having a good relationship between the children as classmates. The present study was carried out with the approval the Ethical Committee of Psychology of the University of Groningen, and the respective informed consent of the children's parents.

### 2.2. Materials

Two sets of hands-on tasks based on the domain of physics (one on air pressure and another on inclined plane) were used (see Appendix A and B). Each set consisted of 6 tasks: one familiarization task, followed by a sequence of five experimental tasks of increasing complexity. The familiarization task was used only to introduce the structure of the solving problem context to the children, and was not included in the results. Hence, the analysis is focused on the five experimental tasks.

The two task sets are designed to elicit spontaneous experimentation in the children. The goal of the task was explained to the children, and they were asked to find out the best way to reach the goal by selecting from several materials those that they thought would work best. The air pressure tasks included the use of materials such as tubes of different length and shape, syringes of different size, a pump, a balloon, small balls of different materials and valves. The goal in these tasks was based on generating air pressure through the connection of syringes and tubes to produce a particular result such as to raise the plunger of a syringe, to rise a lift until the edge of a wooden wall, to rise a platform to a given mark, to blow up a balloon, to make a ball move inside a tube. The inclined plane tasks included the use of wooden marble tracks consisting of marbles of different size and weight, ramps of different length and texture, and wooden columns of different height to provide different inclinations. For this set of tasks, the goal was based on the features of the ramps in order to make a marble roll very far, to roll a ball until it jumps in a basket or a plastic box, to select a height to roll a marble down-up in two connected ramps until it jumps off the ramps.

The air pressure task set (see Appendix A) had been used in a previous study by van der Steen, Steenbeek and van Geert (2012). Similarly, tasks number 4 and 5 of the inclined plane set (see Appendix B) were used in a previous study by Meindertsma, van Dijk and van Geert (2012). In our study, we have developed the protocols for the two task sets in order to create an open problem-solving context that elicits spontaneous experimentation.

## 2.3. Design and Procedure

The seven dyads were assigned randomly to two task conditions: four dyads [Dyad 1, Dyad 2, Dyad 3, and Dyad 4] to the set of the air pressure tasks and the other 3 dyads [Dyad 5, Dyad 6 and Dyad 7] to the set of inclined plane tasks. We used a longitudinal design of six sessions (during one school year) with two months between sessions. Each session lasted approximately 20 to 25 minutes and was video recorded. For each session, we identified the most complex levels of reasoning for both, implicit and explicit knowledge –(we will use the terms 'action events', AE, and 'verbal events', VE, onwards)– in order to examine its development over time, in addition to exploring whether transfer was present in the children's performance (AE). Within each task condition, there is a familiarization task and a sequence of five experimental tasks (T1 to T5) on which we based our analysis (see appendices A and B). The experimental tasks were presented in an overlapping way as follows: In each session two tasks with increasing complexity were presented. We started with the last task of the previous session followed by a new second task (see figure 1). As a result, over the 6 waves of data collection, we obtained a total of 48 measurements in the air pressure set and 36 measurements in the inclined plane set.

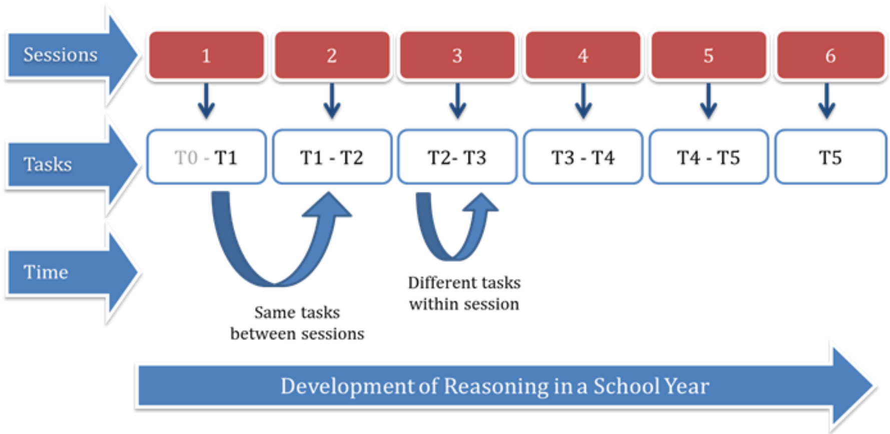



*Figure 1*. Overlapping design of the sequence of tasks. T0= Familiarization task, T1-T5= Experimental tasks.

All tasks were administered by the same experimenter (the first author). For each session, she took the children to a separate room where the task material was set up. The protocol in both conditions included the following five steps: familiarization ("do you know what this is?"), goal-task (e.g. "make a lift to go up with a weight (metal disc) to the edge of this wooden wall"; see Appendix A, tasks n. 2), prediction based on the selection of materials ("what would happen if ..?"), attempts to solve the task and explanation ("What causes this to happen? Do you know why?"). The last three steps (prediction, attempts and explanation) were repeated until the task was solved. If the dyad failed to solve the first task, a second task was introduced after 12 minutes of failed attempts.

## 2.4. Data Coding

For the present study, Fischer's Skill Theory (Fischer, 1980; Fischer & Bidell, 2006; Fischer & Yang, 2002) is used to measure the complexity of children's behaviors. A coding scheme of Meindertsma, van Dijk, Steenbeek and van Geert, (2012) based on this theory, has been used as a guideline to measure the complexity of children's verbalizations (see table 1). In addition, we develop a coding scheme to measure the complexity of children's actions resulting from the spontaneous experimentation (see table 2). The adaptation of the skill theory levels for measuring actions has been less frequently implemented in the literature (e.g. Mascolo, 2013) due to the traditional use of verbalizations as an indicator of reasoning. However, we start from the assumption that both implicit knowledge (AE) and explicit knowledge

(VE) can provide information to have a better understanding of children's reasoning. In this study, we examined the complexity of children's reasoning in the context of a functional level (Bidell & Fischer, 1994), namely as the spontaneous behaviors displayed by the children to solve the task without receiving any teaching or support from the researcher to reach the goal. However, it was possible that the children support each other during collaborative work. As units of behavior, we defined every attempt of the child to solve the tasks as an action event (AE-unit) and every explanation of the child as a verbal event (VE-unit).

Table 1

*Coding Scheme of Skills Levels of Verbal Events (VE)*

| Score | Skill Level | Definition | Examples per Task Set |
|---|---|---|---|
| 1 | Sensorimotor | Describing the object as a whole | *AP:* "Because it is a syringe", *IP:* "Because it is a marble" |
| 2 | Single Representation | Describing attributes of the object | *AP:* "Because it is a big syringe, *IP:* "Because the marble is heavy" |
| 3 | Mapping Representations | Coordinating several attributes (two single representations) | *AP:* "The lift went up because I pushed it" *IP:* "The marble rolls because it is round" |
| 4 | Representational System | Coordinating several relations (two mappings) | *AP:* "This happened because the air goes along the tube and pushes the other syringe" *IP:* "The marble went far away because it is heavy and I put it so high (on the ramp)" |

*Note*. AP= Air pressure task set, IP= Inclined plane task set

Table 2

*Coding Scheme of Skill Levels of Action Events (AE)*

| Score | Solution Type *Skill Level* | Definition | Examples per Task Set |
|---|---|---|---|
| 1 | No Solution<br>*Sensorimotor system* | Isolated action centered in particular element(s) without using functional actions | *AP:* Turning the syringe / Taking a tube and roll it /Turning the syringe barrel<br>*IP:* Holding the marble / Touching the surface of the ramp |
| 2 | Exploration of the task<br>*Single Representations* | Functional actions focused on one or more element(s) without solving the task | *AP:* Pulling out the plunger of the syringe until taking it out/ Trying to connect a the tips of a tube<br>*IP:* Rolling a marble outside the ramp/ Tilting the ramp |
| 3 | Partial Solution<br>*Representational Mapping* | Coordination of functional actions for solving the problem in a partial way | *AP:* Connecting the target syringe with a tube to an incorrect syringe (different size), and pressing the added syringe (as a result the lift overpasses the target goal).<br>*IP:* Using a correct ramp (smooth surface) with and incorrect marble (light) / Using an incorrect ramp (textured) with a correct marble (heavy) (as a result the marble does not roll very far). |
| 4 | Global Solution<br>*Representational System* | Coordination of functional actions with the key elements of the task solving the problem by the discovery of their mechanism | *AP*: Connecting the target syringe with a tube to a syringe of the same size, to make the lift rise up until the goal<br>*IP:* Using a correct ramp (smoothed) and a correct marble (heavy). Then, the marble rolls very far |

*Note.* AP= Air pressure task set, IP= Inclined plane task set. For procedures it is not possible to track the abstract level due to the concrete nature of the actions.

For the VE, we focus only on the explanations. In contrast to the descriptions and predictions, this reasoning skill enables to track the optimal explicit understanding of the causal relations that children establish between the tasks elements. In addition we also coded if every single AE and VE was correct, partially correct or incorrect.

With regard to the examination of transfer, the analysis was based on the individual trajectories of the children (AE) along a sequence of tasks. In this case the verbalizations (VE) were not included for the analysis of transfer, because in several cases children do not provided answers or they reported descriptions of the observed results instead of explaining the relations implied on the tasks. Also it is important to note that in this study, we do not use an instructional phase of learning. Between the consecutive attempts to solve the tasks, the researcher does not point out the similarity between them. Instead, the participants generate their own ideas about each task and explore by themselves how to solve it. Although there was no explicit instruction linking the similarities along the sequence of five tasks increasing in complexity, the provided materials were perceptually (e.g. ramps - column blocks-marbles; syringes- tubes) or functionally similar (e.g. syringes/pump, ramps /curved surface) over tasks. In order to solve the sequence of tasks, children had to use particular materials that we called 'key elements' (see appendices A and B). For instance, in some of the air pressure tasks, the basic mechanism is created with a target syringe (provided in the task) connected with a tube to an added syringe. However, the success of the task depends on selecting the right properties of the materials, such as the size of the added syringe and type of tube. In the case of the inclined plane set, the key elements are related to creating a ramp by selecting column blocks, marbles and the surface of the ramp with the right properties such as a particular height, weight and texture respectively. Therefore, we operationalized transfer in terms of the 'key elements' required to succeed in the tasks.

Our interest in using these exploratory tasks was to provide the children a context to discover the variance and regularities of the mechanisms (i.e. air pressure or inclined plane task set), as is often the case in naturalistic conditions. As we have mentioned before, we assumed that transfer is the result of the self-organization process of the child and the task. Therefore, with the increasing difficulty of the tasks, the ability to transfer needs a re-configuration of the tasks key elements. In each new task, children should go beyond the perceptual appearance of the elements that allowed them to solve the previous task (attributes of the key elements) to focusing on their structural functionality to discover the new key elements. For instance, children should discover that a particular size of a syringe does not lead to the solution but the amount of air that it can generate does.

The inter-rater reliability was established by comparing independently coded videos by the first author and by a second trained rater. For the air pressure tasks, 75% (behaviors of three dyads for the six waves, equivalent to 18 complete videotapes) of the data was coded by both raters and 66% (behavior of two dyads for the six waves, equivalent to 12 complete videotapes) for the inclined plane task. For each task set, the unit of actions was defined as each single attempt to solve the task, and for verbalizations the unit was every explanation provided by the child. The kappa (Table 3) was 'almost perfect' for the main categories (range of 0.81–0.99; see Viera & Garrett, 2005). In addition, the probability that they were based on chance is in all cases very small (see p-values).

Table 3
*Kappa Calculation for all Categories with Bonferroni Correction*

| Task | Kappa Skill Level | | | | Kappa Correct-Incorrect | | | |
|---|---|---|---|---|---|---|---|---|
| | AE | P-value | VE | P-value | AE | P-value | VE | P-value |
| Air Pressure | .971 | $p$=.001 | .928 | $p$=.001 | .952 | $p$=.001 | .980 | $p$=.001 |
| Inclined Plane | .929 | $p$=.001 | .942 | $p$=.001 | .932 | $p$=.001 | .878 | $p$=.001 |

*Note.* AE= Action events, VE= Verbal events.

## 2.5. Data Analysis

As mentioned before, we focused our analysis on the most complex behaviors of AE and VE displayed by the children in each session. For this reason, we have selected the AE and VE with the highest complexity level among all attempts for each child in each single task. We conducted different analyses as follows:

(1) A global overview of children's performance is provided by the use of descriptive statistics as the average of skill levels, transfer, correct/partial/incorrect solutions, in the different task sets and different conditions (AE, VE) over time and attempts.

(2) For a temporal approach of the data, time-series graphs (group and intra-individual) were plotted in order to describe the changes in children's performance over the sessions.

(3) The analysis of transfer was based on the use of key elements to solve the tasks (AE) (see Appendices A and B). Because the tasks were different between and within conditions (air pressure, inclined plane), we identified the number of relations that children needed to make in order to solve the task. After this, we established a percentage of 'successful performance' for each single task based on the number of key elements and relationships that were used between tasks. Thus, transfer is operationalized as is the use of particular key elements to solve the tasks. These key elements have different perceptual attributes but a similar functionality. Additionally, for each task set we calculated *the changes in the proportion of transfer* (use of key elements) from one task to the next one. This procedure consisted of a descriptive quantification of the percentage of children that showed an e*qual*, *decrease,* or increase of transfer along the sequence of tasks (T) of air pressure and inclined plane (e.g. T1- T2, T2-T3, T3-T4, T4-T5).

(4) Finally, in order to determine the probability that the observations are based on chance, Monte Carlo analysis (Good, 1999) was performed. This method consists of running multiple trials based on shuffled data (10000 random permutations) in order to determine whether the empirical results are different from what would be expected if the data correspond with a random distribution. The following tests were performed:

- *To test the differences of the number of attempts / highest complex attempt between and within task sets*: we first shuffled the total average of attempts along the six sessions, for AE and VE respectively. After this, we compared –within and between task sets- the average of attempts of the empirical data with the shuffled data. We repeated this procedure to test the attempt in which the child reached the highest complexity level in each session attempt for AE and VE.

- To test the *differences of the skill levels of AE and VE between and within task sets:* for each task we shuffled the scores of optimal skill levels, computing the group average of AE and VE. Second, we compared –between task sets- the empirical group average of skill levels of AE and VE with the respective shuffled data. The same procedure was repeated at the intra-individual level (individual average of skill levels of AE and VE).

- *To test the developmental change of the skill levels of AE / VE between task sets*: we used the shuffled data that was used for the skill levels test, but, in this case, we computed and shuffled the average increase (linear slope) of AE and

VE respectively. After this, we computed the difference in the slopes of AE and VE between tasks in the empirical and shuffled data. Finally, we compared this difference between the slopes in the empirical and shuffled data respectively. This procedure was repeated at the intra-individual level.

- To test the *variability of the skill level trajectories*: we calculated the amount of variability in the trajectories of AE and VE for both task sets. First, within task condition, we averaged the skill levels of the individual trajectories for each session. Second, independently for AE and VE, we added the differences or changes of this trajectory from one session to the next. For instance, if the average skill levels from the first session to the third one are 3, 4 and 2, the resulting difference between sessions is 3. Third, we shuffled the individual trajectories of each task condition (air pressure and inclined plane), and we repeated the procedures for calculating the average and the estimation of variability for each task set. Finally, within tasks, for empirical and shuffled data, we subtracted the variability of AE with the variability of VE and estimated the probably that this difference was based on chance.

- *To test the level of success in transfer and its development for each task condition and child*: First, per child, we shuffled the average number of key elements used to solve the task along the six sessions. Second, we computed the total average of success (key elements) and the slope of transfer for each task. Finally, we compared the total average of success and the slope of the empirical data with the respective shuffled data. The same procedure was carried out at the intra-individual level to examine the variations in the development of transfer.

The Monte Carlo analyses were carried out in Microsoft Excel in combination with PopTools (version 3.2). Being aware of the controversial use of a specific predetermined alpha to determine significance (see Gigerenzer, 2004; Schneider, 2015), we only report the exact p-values which indicate the probability that the finding is based on chance with the procedures described above. In addition, we present a descriptive characterization for the intra-individual trajectories.

## 3. Results

First, we provide a global overview of the children's performances as the total number of attempts and the attempt in which the most complex AE and VE appeared. Based on the most complex level of AE and VE, (a) we inspect the developmental change over the entire measurement period and finally (b) the spontaneous transfer (AE) of the two task sets. In both cases, we also provide a close examination of the intra-individual trajectories.

### 3.1. Overview of Children's Reasoning in Actions (AE) and Verbalizations (VE)

Table 4 shows the group average of the number of attempts per task set (AE and VE), as well as the average of the number of attempts the children needed to reach their highest performance. Here we observe that for both tasks, children used a higher amount of AE attempts compared with the VE attempts of the same task set. In addition, we found that children working on the air pressure set needed more attempts before reaching the highest level of AE and less attempts to reach their highest level on VE than the children working on the inclined plane task. A Monte Carlo test revealed that the probability that these differences are based on chance were very small, both for the total number of attempts ($p$=.001) as for the number of attempts they need to reach their highest performance ($p$=.001). This indicates that the air pressure set was probably more complex, requiring more attempts to be solved than the other task set.

Table 4

*Average of number of attempts performed by the children in the two task sets and the corresponding average of the number of attempts with the highest skill level for AE and VE*

| Task Set | Average of Total Attempts | | | Attempt Average with the Highest Skill Level | | |
|---|---|---|---|---|---|---|
| | Action (AE) | Verbal (VE) | P- value[1] | Actions (AE) | Verbal (VE) | P- value[1] |
| Air Pressure | 9.6 | 2.3 | $p$=.001 | 7.4 | 1.3 | $p$=.001 |
| Inclined Plane | 3.4 | 5.7 | $p$=.001 | 2.2 | 2.7 | $p$=.001 |
| P- value[2] | $p$=.001 | $p$=.001 | | $p$=.001 | $p$=.001 | |

*Note*. P-value[1] = Comparison *within* tasks, P- value[2]= Comparison *between* tasks

## 3.2. Overview of the Complexity in Children's Actions (AE) and Verbalizations (VE)

Figure 2 shows the group trajectories, presented in terms of optimal skill level per session. Here, we observe that on average, children showed higher skill levels for AE than for VE within each set of tasks, consistently over all sessions. Moreover, there are clear differences in the trajectories of these two task sets.

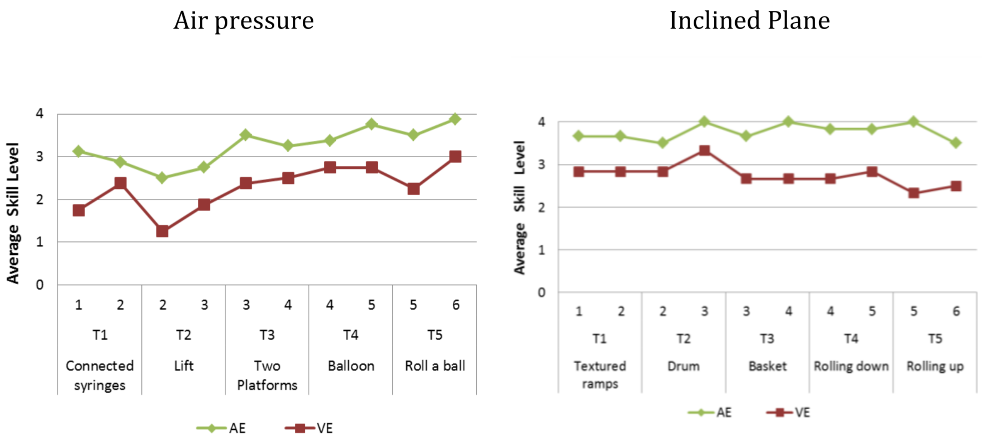



*Figure* 2. Average of children's optimal skill level of actions (AE) and verbalizations (VE) in the two task sets. T1 to T5= sequence of tasks presented in overlapping sessions (e.g. T1 in sessions 1-2, T2 in sessions 2-3, etc.).

In the air pressure set, the skill levels of AE and VE show a global increase in complexity level across time. In particular, the AE were characterized mainly by partial solutions (skill level 3- representational mapping, M= 3.25) and less by correct solutions. For instance, children with a partial comprehension were focusing on building a mechanism (e.g. connected syringes), but they fail by omitting the structural functionality of the materials. In contrast, children with a better comprehension of the tasks found out that the structural characteristics of the elements matter to reach the goal (e.g. the syringe's size is related with the amount of air pressure that can produce). On the other hand, children's explanations (VE) were mainly focused on single attributes of the task elements (skill levels 2 - single representations), (e.g. "Because of the air!", "Because it is big [the syringe]") and causal relations between the task elements (skill level 3- representational mapping) (e.g. "Because when you push here [one of the syringes], it blows", "Because the air... goes so big and there was so much air").

Regarding the inclined plane set, the skill levels of AE and VE seem to be relatively stable over sessions. The AE were mainly characterized by the combination of partial and correct solutions (skill level 3 (representational mapping) and skill level 4 (representational system), M=3.77) in which children partially or totally build and solve the task (i.e. mechanism of air pressure or mechanism of inclined plane). For example, children with a partial comprehension of the task (skill level 3) focused their performance on building a ramp, paying less attention to other characteristics as the type of marble, the relation of the height of the ramp with the goal to be reached. In contrast, children who successfully solve the task (skill level 4) were able to integrate these particular attributes of the materials in their performance. In contrast, the VE were based on the descriptions of the materials (skill level 2-single representations) (e.g. "Because that one [ramp] is not so high", "Because it [marble] is too light") and the partial explanations of the mechanism of the task (skill level 3- representational mapping) (e.g. "Because it [marble] is so small that it goes really fast"), rather than on the complete explanation of the mechanism of the task.

In addition, the Monte Carlo analysis showed that the skill levels of AE were higher than the skill levels of VE for both the air pressure set ($M_{AE}$=3.25, $M_{VE}$=2.29), ($p$=.001) and the inclined plane set ($M_{AE}$=3.77, $M_{VE}$=2.75) ($p$=.001). However, the analysis revealed that there are no indications that there is a difference *between* the two tasks, neither for AE ($p$=.601) nor for VE ($p$=.406). Furthermore, the analysis of the increase of skill levels (slope) in both task sets, showed an increase of complexity for the VE in the inclined plane set over the air pressure set ($p$=.001), and for the AE in the air pressure set compared with the inclined plane ($p$=.030). The variability of the group trajectories of AE and VE in both tasks sets seemed to show slightly higher changes in the trajectories of VE (air pressure set = 5, inclined plane set = 2) than for AE (air pressure set = 3, inclined plane set= 0). However, the Monte Carlo analysis indicate that the probability that these differences are based on chance are relatively large (air pressure, p=.362; inclined plane, p=.360). This means that the variability of the average trajectories is roughly the same for VE as for AE.

### 3.3. Intra-individual Trajectories of the Complexity of Children's Reasoning

Looking at the group data, we have seen a clear pattern on the two trajectories of reasoning such as AE and VE. Nevertheless, in order to gain more information about the emergence of the children’s reasoning in real-time, we also have to zoom-in and inspect the respective intra-individual trajectories of reasoning. Figures 3 and 4 present the intra-individual trajectories of AE and VE for the air pressure and inclined plane. Here, we observe that the individual trajectories are mainly characterized by idiosyncrasy. In some of the individual trajectories we can observe the higher levels of AE in relation to the VE, while in other cases, an equal level between the level of AE and VE is observed.

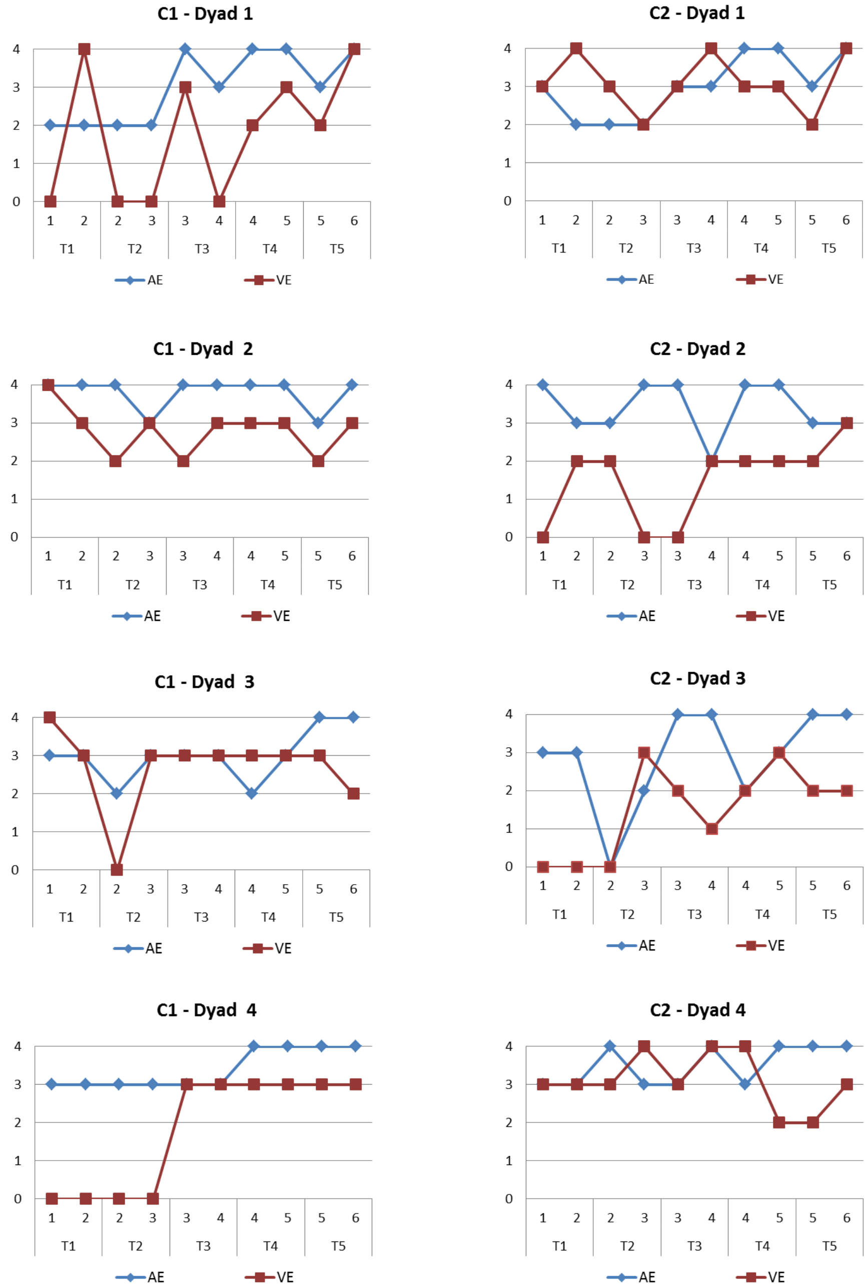



*Figure* 3. Skill levels of the children's action (AE) and verbalizations (VE) in the air pressure task set (T1 –T5). Members of the dyad: C1= child 1. C2= Child 2. See p-values - table 5

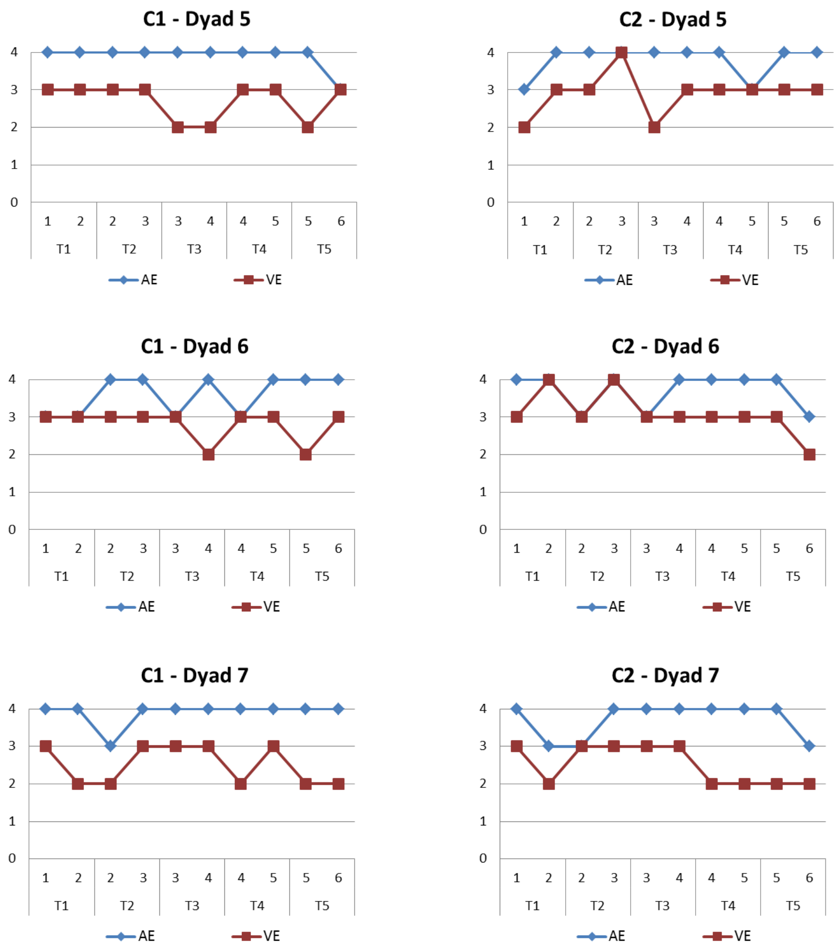



*Figure 4*. Knowledge level of the children's actions (AE) and verbalizations (VE) in the Inclined plane task set (T1 –T5). Members of the dyad: C1= child 1. C2= Child 2. See p-values in Table 5.

The differences between the AE and the VE were tested for each individual with the Monte Carlo analysis. The results from the air pressure set (see Figure 3 and Table 5), showed 5 out of 8 children in the air pressure task and all 6 children of the inclined plane task, had higher skill levels for AE than for VE (see Figures 3 and 4, and Table 5). Moreover, table 5 shows that despite the idiosyncrasy of the children's performance, there were no differences between VE and AE in the intra-individual increase (slope) of the skill levels neither for air pressure or inclined plane.

Table 5

*P-values of the Monte Carlo Test comparing by children the difference between skills levels and slopes for AE and VE*

| Task Condition | Dyad | Child | P-value *Skill levels* (AE vs. VE) | P-value *Slope* (AE vs. VE) |
|---|---|---|---|---|
| Air Pressure Set | Dyad 1 | C1 | .046 | .509 |
| | | C2 | .728 | .046 |
| | Dyad 2 | C1 | .002 | .393 |
| | | C2 | .001 | .880 |
| | Dyad 3 | C1 | .337 | .202 |
| | | C2 | .018 | .638 |
| | Dyad 4 | C1 | .004 | .909 |
| | | C2 | .154 | .060 |
| Inclined Plane Set | Dyad 5 | C1 | .001 | .537 |
| | | C2 | .002 | .785 |
| | Dyad 6 | C1 | .003 | .099 |
| | | C2 | .024 | .144 |
| | Dyad 7 | C1 | .001 | .311 |
| | | C2 | .001 | .180 |

### 3.4. Overview of Transfer of scientific reasoning in children's performance (AE)

In figure 5, we present the average percentage of successful transfer (use of key elements of the task) for each session in both tasks. On average, children on the inclined plane set showed a high-stable percentage of transfer between tasks (ranges from 83.3% to 91.7%). In contrast, children working in the air pressure set presented a medium-high percentage of transfer between tasks (ranges from 58.3% to 72.5%). On average, there is an increase in their rate of transfer between session 2 and 3 (54.2% to 87.5%), after which a slight decrease seems to set in. This suggests that in the air pressure task the transfer is affected by the task series (complexity or content). This is not the case for the inclined plane tasks, were children's transfer appear relatively stable along the different tasks.

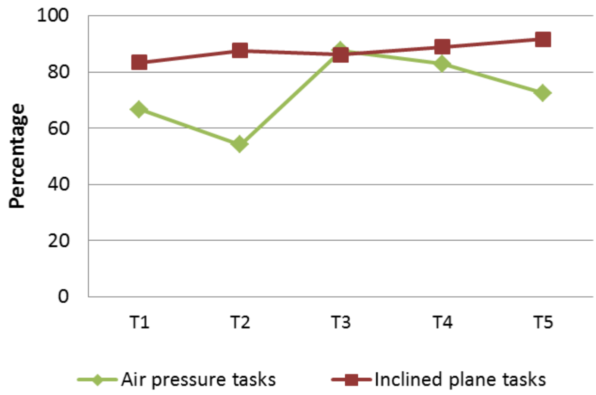



*Figure 5*. Percentage of children's success (in terms of the selection of key materials) solving the task sets of air pressure and inclined plane.

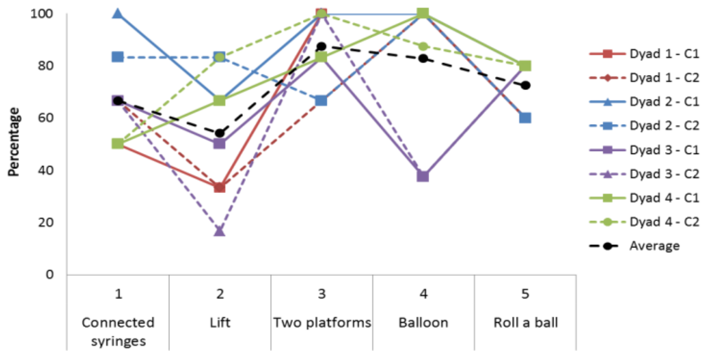
A Monte Carlo analysis revealed that the proportion of transfer (key elements) for the inclined plane set (M= 87.5) seems to be larger than in the air pressure task set (M=72.7; *p*=.020). There are no indications that the increase (slope) of transfer is larger for one of the two task sets (*p*=.323).

### 3.5. Intra-individual Trajectories of Transfer of Scientific Reasoning (AE)

For a closer inspection of the development of transfer in real-time, figure 6 shows the individual trajectories, representing the percentage of transfer of AE in both task sets. In relation to the group data of transfer, the intra-individual data added information about the differences within and between tasks. The average of transfer in the air pressure set showed medium-high values ($M_{tasks}$= 71.1%) and high variability from the beginning until the end of the children's trajectories, shaped as progressions and regressions (scalloping). In contrast, the average of transfer in the inclined plane set, presented a high average of transfer ($M_{tasks}$= 87.5%) and reduced variability within and between the children's trajectories.

In general, the group and the intra-individual data showed that the ability of children to transfer structural information from one task to the next varies over time. For both task sets, transfer does not necessarily decrease with the complexity of the tasks, but shows a constant degree of fluctuation between the sequences of tasks.

Air Pressure task set




Inclined plane task set

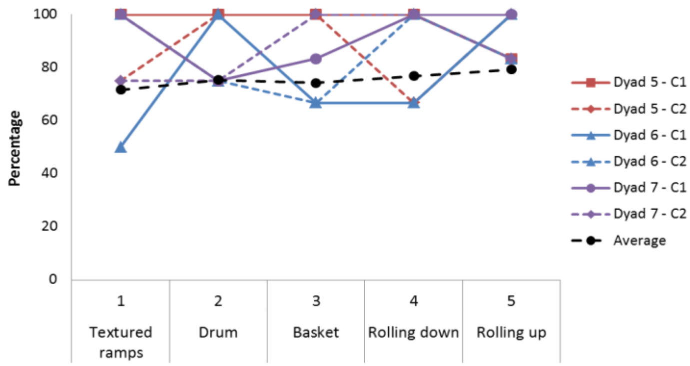



*Figure* 6. Individual percentage of transfer (selection of key materials) used by the children in the task sets of air pressure and inclined plane. C1= child 1. C2= child 2.

For each task condition, the changes of transfer from one task to the next are presented in figure 7. We can observe that between a pair of tasks, the proportion of transfer is not stable (i.e. use of equal amount of key elements), but instead shows fluctuations by increasing or decreasing the amount of transferred elements. In the air pressure set, two salient changes are observed. One corresponds to the decrease in the children's performance of transfer at the beginning (T1-T2) and at the end (T4-T5) of the sequence of tasks. The other corresponds to the increase of the proportion of transfer (T2-T3). In the inclined plane set, different characteristics of transfer are observed. For instance, in half of the tasks, the children's performances are equally distributed between equal, decrease and increase of transfer (i.e. use of key elements).

In contrast, in the second half of the tasks, children's transfer is more consistent, showing a combination of predominant performances as equal/increase of transfer (T3-T4 = 50% and 33.3% respectively) and decrease/increase of transfer (T4-T5= 50% and 33.3% respectively). These results show a nonlinear process of transfer, because across all sessions there are frequent increases, but also decreases in the amount of information transferred from one task to the next one.

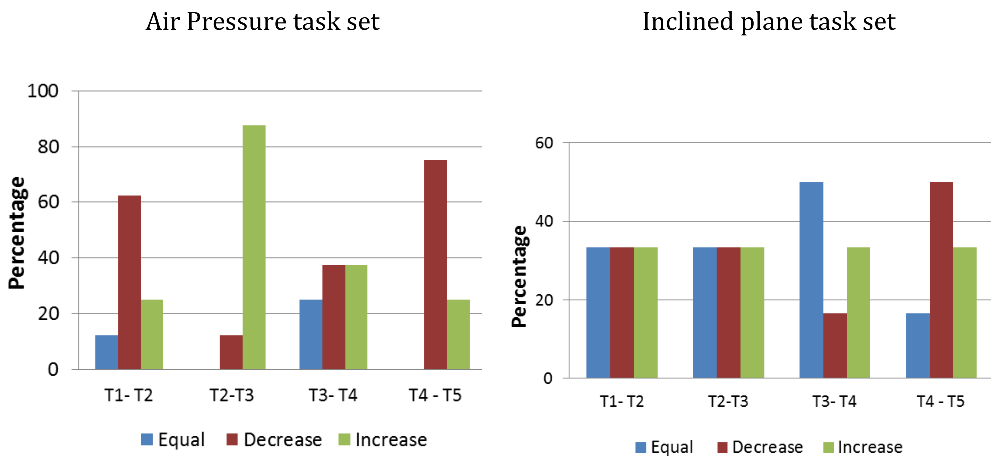



*Figure* 7. Proportion of the transfer showed by the children in each task set from one task to the next one. T1 to T5 = Task 1 to Task 5.

## 4. Discussion

The aim of this study was twofold. The first was to examine the developing complexity of young school children's reasoning in action events (AE) as implicit knowledge and verbal events (VE) or as explicit knowledge during a spontaneous experimentation in two sets of hands-on tasks. The second aim was to inspect whether these children transfer the information in their actions gained in a task to solve a new one. With regard to the first aim, we looked at how children performed in each task set (air pressure and inclined plane) by analyzing their optimal skill level of AE and VE in each session. We found that on average the 5-year-olds showed higher complexity levels of implicit knowledge (AE) than of explicit knowledge (VE) in both task sets, across all sessions. The fact that children find it more difficult to verbalize their understanding of the task than showing it through their actions, support the assumption that implicit knowledge usually precedes explicit knowledge (Reber, 1993; Siegler & Stern, 1998; Sun, Merrill & Peterson, 2001). However, from a more dynamic perspective, it may also mean that within task sets, AE and VE require a different process of soft-assembly. From a complex dynamic systems perspective, it can be seen as the implicit knowledge (AE) following a different process of short-term emergence than the explicit knowledge (VE) does.

Furthermore, our results showed task differences for both the optimal skill levels of AE and VE. In the inclined plane set, children need fewer attempts to solve the task (AE), but their explanations (VE) are less complex. In contrast, for the air pressure task, children need more attempts and often solve the task only partially (AE), but show relatively similar skill levels of explanations (VE) than the children working on the inclined plane. Possible explanations of the difference in the number of attempts include the children's familiarity with the task's affordances (Greeno, 1994; Meindertsma, van Dijk & van Geert, 2013) and the respective gain of information obtained as a result of the inconsistency with prior knowledge (Legare, 2014). In our study, the air pressure task probably shows more inconsistency with prior knowledge than the inclined plane set, due to the unfamiliarity of the children with the task structure (e.g. syringes and tubes vs. marble tracks) and goals (e.g. to rise a lift vs. to make a marble roll) of the tasks. For this reason, the knowledge gain may be greater in the air pressure set in which children need to learn from their failed attempts to adjust their actions compared to the inclined plane set, where the children's familiarity with the task facilitates its solution. Subsequently, in the unfamiliar task, we can speculate that the children's knowledge gain through their procedures (AE) may have had a positive influence on the complexity level of their explanations (relatively high skill levels of VE). The results suggest a different process of soft-assembly between implicit and explicit knowledge, due to a prevalence of the actions (AE) over the sequence of verbalizations (VE). Although it seems that this difference between AE and VE is not task dependent, the characteristics of the tasks –such as the familiarity with the task affordances and the consistency with children's prior knowledge– seems to have influenced the complexity levels of implicit and explicit knowledge reached by the children.

When inspecting the average development, the difference between VE and AE is relatively constant across the entire measurement period. However, a closer examination of the intra-individual trajectories showed more differences over time (e.g. scalloping, Fischer, 2008). In addition, the intra-individual variability in the children's reasoning adds evidence to previous studies about the non-linearity of scientific reasoning. For instance, a microgenetic study by Meindertsma, van Dijk, Steenbeek and van Geert, (2012) on predictions and explanations in a floating and sinking task indicated that 5-year-olds presented high inter- and intra-individual variability in their complexity levels and content of their explanations. Another example from the longitudinal study of van der Steen, Steenbeek & van Geert (2014) shows constant fluctuations in the complexity of a 4-year-old child's understanding of air pressure. This indicates that individual development is in many cases much more variable than the group results suggest. Therefore, our data add plausibility to the fact that variability is a property of the intrinsic dynamics.

With regard to transfer, the children's performance showed high variability, which is a typical feature of learning in general (e.g. Siegler, 2007). Based on the variability of the empirical data, we might re-consider the concept of transfer as an automatic exportation of information in favor of a more dynamic approach in which transfer is an ongoing process of the interplaying of the history of knowledge of the learner and the affordances of the task. In daily life, learning does not occur as the accumulation of concrete events or experiences, but as a continuous flow of knowledge in transformation. In contrast to traditional approaches, we have measured transfer not in a single measurement or in terms of either success or failure, but in repeated trials, examining the levels of success of the children's actions. The focus was on the use of key elements of the tasks as indicator of the children's comprehension of the mechanisms of air pressure and inclined plane. The overview results of children's success revealed that between task sessions, children were able to transfer the use of most of the key elements to build the respective mechanisms of air pressure and inclined plane. Overall, the main characteristic of trajectories of transfer is the nonlinearity observed between tasks but also along the sessions.

Although these results show that children identified the structure of each task (i.e. mechanism of air pressure and inclined plane) and tried to re-create it in a new task, it is difficult to establish whether and how the children's performance is influenced by their experience with the previous tasks (e.g. task 1-task 2; task 2-task 3…). For instance, the question is whether the children would have achieved the same proportion of success in the second task if they had not worked with the first task at all. Consequently, it can be argued that one of the main limitations of the study is the lack of a 'control group' in order to test the children's success solving a particular task, without the sequence of tasks. Despite this possible limitation, we can still consider two aspects: First, with the increasing complexity of the tasks, if children were not transferring any kind of information, their performance would probably have shown a decline as they reached the most complex tasks. A possible explanation of children's success from task to task is the use of the 'constraint violation' mechanism (Nokes, 2009). As has been mentioned, learners using this mechanism of transfer go beyond the perceptual information in order to 'generate-evaluate-revise' the actions to reach the goal. Nevertheless, from a dynamic approach, transfer as a process that emerges in real-time, 'here' and 'now', is highly likely to involve a soft-assembly process between the learner and the task (e.g. affordances of the tasks and the previous performance history of the learner). Therefore, the adjustment of the children's knowledge to solve each new task is a constant dynamic process in which previous knowledge and new knowledge are always interacting elements.

Second, if children would only base their performance on the analogic mechanism of transfer, their performances would have succeeded only for the initial tasks, because the perceptual structure of later tasks was totally different (see appendices A and B). The results show that the percentage of success does not decline, which does not mean that children do not use analogical mechanisms, but that they are adjusting the relations discovered in the previous task. In addition, what we call transfer is in fact a part of a dynamic, soft assembly process of the history-dependent solving of a series of problems. Therefore, we suggest that the nature of transfer is not static but dynamic, a process that is continuously updated with the previous knowledge of the learners, the characteristics and the demand of the task, and obtaining a variety of performances with different levels of accuracy.

The intra-individual variability of the trajectories of transfer reflects the consistency of the children's understanding of the key elements of the tasks. We find relative medium-high (air pressure) and high (inclined plane) proportion of success on the tasks. If the children's performance would be caused only by a trial-error strategy, we would not have found in all children the consistent trends of success across time. However, it is important to notice that the intra-individual variability of the trajectories in some cases resemble the group data, but in others not. This variability indicates that reaching a higher level of transfer does not imply a subsequent stability.

The combination of the analysis of group trends and intra-individual measurement allows us to show the underlying dynamics of implicit and explicit knowledge at a point in time in which the complexity of reasoning is expected to show a transition from sensorimotor to representational levels (Fischer & Bidell, 2006). In this study we have captured all actions and verbalizations of the child in a specific task in order to identify the optimal level of reasoning and follow its changes across time. This would not be possible with a single-shot measurement. Furthermore, in our study, the complexity of reasoning is described on the basis of the children's behaviors in a naturalistic task context. The skill level perspective focuses on the complexity of the cognitive functioning and not on the dichotomy of success and failure.

In summary, this study has shown that the spontaneous experimentation of 5-year-olds~~-~~working in these task series~~-~~, consistently show a dominance of AE over VE. However, the individual and task differences should not be underestimated. The short- term processes of reasoning and transfer show nonlinearity and variability, which fit with a complex dynamic systems interpretation of these processes. Moreover, it seems that the dynamic loop of exploration and explanation not only works as a mechanism for generating and testing hypotheses (Legare, 2012, 2014), but also influences the levels of reasoning of the implicit and explicit knowledge.

An implication of our findings is that teachers should be aware that knowledge is not only present in verbalizations but also in actions (Goldin-Meadow, 2014). In activities combining 'doing' and 'saying', children have more opportunities for evaluating the inconsistency of their previous knowledge and are able to display relatively complex levels of reasoning.  However, the difference between the verbal and nonverbal ways of reasoning is not present for all children and not at every moment. Therefore, teachers should be aware of individual and temporal differences, as is the case for many aspects of learning.

We acknowledge that from a traditional perspective, the small number of participants can be seen as a limitation. However, from a process-oriented view, using a small sample enables us to zoom in on the time serial process of change. Another limitation is that in the current analyses we did not look at the moment-to-moment variability within sessions. The reason for this is that the number of attempts was very different in each session, which complicated a straightforward interpretation of the group data. Future research with a defined range of attempts would be suited to explore the real time variability of children´s behaviors. It should also be noted that this study focused on children's individual spontaneous performance when being paired up with a peer. However, it is known that children can show more complex performances when they receive the support of an adult (Rappolt-Schlichtmann, Tenenbaum, Koepke & Fischer, 2007). Therefore, in an educational setting, teachers can combine scaffolding techniques (e.g. using follow-up questions) with an instruction based on the framework of the "empirical cycle" of scientific reasoning in order to provide a rich context to foster learning and transfer of scientific concepts (Wetzels, Steenbeek & van Geert, 2015).

Finally, with regard to the research on transfer, a process approach is highly suited to provide a better understanding of the nature of transfer as a mechanism of learning because it replaces the traditional concept of transfer for a notion of history dependence of the thought processes over the medium-term of repeated sequences of tasks. We have shown that 5-year-olds are able to show spontaneous transfer in a domain of knowledge in which they have not received instructions. The characteristic of the tasks seems to support the 'constraint violation' mechanism of transfer suggested by Nokes (2009) as a procedure that allows learners to transform their understanding. In this sense, transfer should be thought of as a continuous process instead of a local segmented procedure from one task to the next. In conclusion, both, transfer and reasoning are considered to be essential for a better understanding of learning, and, at the same time can provide a basis to foster the development of scientific skills in early childhood education.